

Capacity Achieving Peak Power Limited Probability Measures: Sufficient Conditions for Finite Discreteness

Vincenzo Zambianchi, *Student Member, IEEE*, Enrico Paolini, *Member, IEEE*,
and Davide Dardari, *Senior Member, IEEE*

Abstract

The problem of capacity achieving (optimal) input probability measure (p.m.) has been widely investigated for several channel models with constrained inputs. So far, no outstanding generalizations have been derived. This paper does a forward step in this direction, by introducing a set of new requirements, for the class of real scalar conditional output p.m.'s, under which the optimal input p.m. is shown to be discrete with a finite number of probability mass points when peak power limited.

Index Terms

Channel capacity, discrete input, conditional output probability measure, real scalar channels.

I. INTRODUCTION

In recent years, a great interest has been rising in what can be called discrete input channel modeling. This theory takes its first steps from the study of classical (Gaussian) additive noise channels under input constraints. The class of channels with input limitations is important from a practical point of view since feasible systems do always have to deal with input constraints: Peak and average power are necessarily bounded. The first works in this field were the ones by Smith back in the 70's [1], [2]: He made forward steps with respect to Shannon's work [3]

The authors are with the Department of Electrical, Electronics and Information Engineering "G. Marconi" (DEI), University of Bologna, via Venezia 52, Cesena (FC) 47521, Italy. E-mail: {vincenzo.zambianchi, davide.dardari, e.paolini}@unibo.it. This work has been supported by the GRETA PRIN Project.

considering an additive Gaussian noise channel in which the input is either peak or both peak and average power constrained. He discovered that, under both constraints, the capacity achieving input p.m. is discrete with a finite number of probability mass points. This kind of p.m.'s will be referred to as *finitely discrete* throughout this paper. Smith's result was of notable importance since continuous inputs are not feasible in practice and have to be approximated with finitely discrete inputs.

The finitely discrete feature was demonstrated to be the exact solution for the capacity achieving input p.m. in the constrained additive scalar Gaussian noise channel model. This paved the way to several subsequent studies that, more recently, explored the finite discreteness of capacity achieving input p.m.'s for other input constrained channel models, presenting quite disparate characteristics. Among them we cite [4] and [5], which inspired further works such as [6] and [7]. Concerning the two last mentioned works, the former presents conditions on the p.m. of an additive scalar channel noise, that are sufficient for the optimal bounded input p.m. to have a finitely discrete support. The latter demonstrates that such a support is sparse (see [7] for definition) when the channel conditional output p.m., possibly not scalar, is Gaussian distributed. Subsequent works exploited the finitely discrete nature of the input p.m. in some specific cases (e.g., [8], [9]) but no further generalizations have been developed to the authors' knowledge.

In this paper, we consider a wide real scalar channel model and provide sufficient conditions on the conditional output p.m. for the peak power limited capacity achieving input p.m. to be finitely discrete. We establish this result without indicating any particular type of conditional output p.m. nor any particular kind of the channel input-output law. Moreover, we prove that several peak power constrained additive channels as well as the peak power constrained Rayleigh fading channel fall in the developed framework as particular cases, whereas so far they have always been regarded as two distinct categories, necessitating different mathematical treatments. In this respect, the presented conditions extend the theory of peak power limited real input scalar channels.

The contribution is organized as follows. In Section II all necessary notation and definitions are introduced, while in Section III our main result is stated. This result is gradually proved in Sections IV, V, and VI. Some hints about uniqueness of the capacity achieving input p.m. are provided in Section VII. The above mentioned examples are analysed in Section VIII, while conclusions are drawn in Section IX. Ancillary results necessary for the proof of the main theorem

are deferred to Appendices A, B, and D while Appendix C provides some deeper explanations concerning the earlier discussed examples.

II. NOTATION AND EARLY DEFINITIONS

In this section we present our notation and definitions coherently with the ones given by previous authors [2], [7].

Throughout this paper, Y and X represent the real scalar channel output and input random variables (RVs), respectively. We denote by $F(x)$ the input cumulative distribution function (c.d.f.), by $p_X(x)$ the input p.m., and by $p_{Y|X}(y|x)$ the conditional output p.m. The input RV X is assumed to take values in the set \mathbb{S} , with \mathcal{P} being the ensemble of possible p.m.'s defined on that set. The corresponding class of c.d.f.'s is denoted by \mathcal{F} . We have

$$p_Y(y) = \int_{\mathbb{S}} p_{Y|X}(y|x)p_X(x)dx = \int_{\mathbb{S}} p_{Y|X}(y|x)dF(x) = p_Y(y; p_X) = p_Y(y; F) \quad (1)$$

where we make explicit the dependence on $p_X(x)$ of the output p.m. $p_Y(\cdot)$.¹

Channel capacity is the supremum over the input p.m. of the mutual information functional [10]

$$I(X; Y) = \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{p_Y(y; F)} dF(x) dy = I(F) \quad (2)$$

where $\log(\cdot)$ denotes the base-2 logarithm.² Since only meaningless channel structure have zero capacity, we will assume channel capacity to be strictly positive and we will denote the capacity achieving (hence optimal) input p.m. by $p_{X_0}(x)$. The mutual information functional can be further developed as

$$I(F) = H(F) - D(F) \quad (3)$$

where

$$H(F) \triangleq - \int_{\mathbb{R}} p_Y(y; F) \log p_Y(y; F) dy$$

and

$$D(F) \triangleq - \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \log p_{Y|X}(y|x) dF(x) dy.$$

¹Here, and throughout the whole paper, one of the two equivalent formulations with $p_X(x)$ or $F(x)$ will be freely used as appropriately needed.

²In contrast, $\ln(\cdot)$ will denote the natural logarithm.

We can note how $D(F)$ depends in general on the input c.d.f., as opposed to what happens for an additive Gaussian channel (Smith's model, [2]).

We also define the marginal information density and the marginal entropy density as

$$i(x; F) \triangleq \int_{\mathbb{R}} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{p_Y(y; F)} dy$$

and

$$h(x; F) \triangleq - \int_{\mathbb{R}} p_{Y|X}(y|x) \log p_Y(y; F) dy$$

respectively. These two densities are related as

$$i(x; F) = h(x; F) - d(x)$$

where

$$d(x) \triangleq - \int_{\mathbb{R}} p_{Y|X}(y|x) \log p_{Y|X}(y|x) dy.$$

It is straightforward to show that the following three statements also hold:

$$I(F) = \int_{\mathbb{S}} i(x; F) dF(x) \tag{4}$$

$$H(F) = \int_{\mathbb{S}} h(x; F) dF(x) \tag{5}$$

and

$$D(F) = \int_{\mathbb{S}} d(x) dF(x). \tag{6}$$

In this paper, (4), (5), and (6) are well-defined since $h(\cdot)$, $i(\cdot)$, and $d(\cdot)$ are finitely bounded under the conditions enunciated in Section III, as proven in Appendix A.

III. FRAMEWORK SET UP AND MAIN RESULT

We consider a memoryless real scalar channel governed by a general input-output relationship in the form

$$Y = f(X, \underline{\Theta}) \tag{7}$$

where X is the input RV and $\underline{\Theta}$ a vector of nuisance parameters. We do not impose further conditions on the input-output channel law $f(\cdot)$, which may be linear or nonlinear, additive in noise or multiplicative or both, with independent or correlated noises.

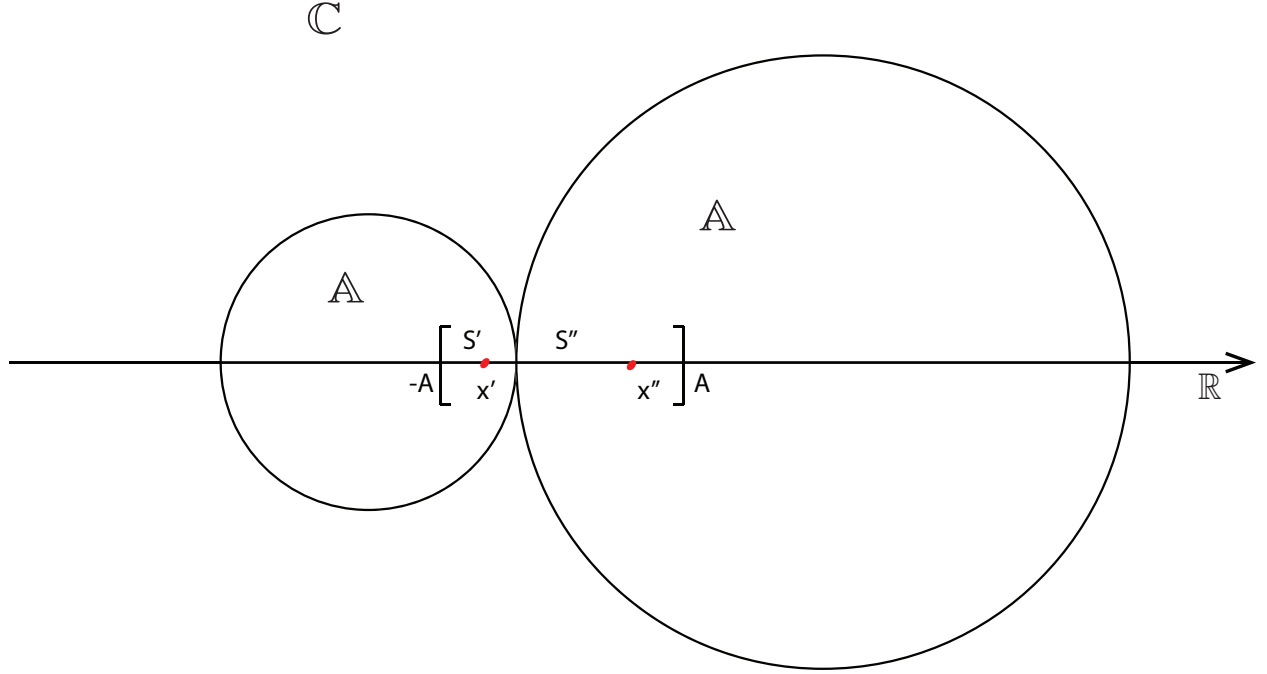


Fig. 1. Pictorial representation of the set $\mathbb{S} = [-A, A] \cap \mathbb{A}$ on which the input RV takes its values.

Throughout the paper, we consider a peak power constrained input RV X taking values in the bounded set (see Fig. 1)

$$\mathbb{S} = [-A, A] \cap \mathbb{A}$$

where $[-A, A]$ is the compact real interval of radius A and \mathbb{A} represents an open subset of the complex extended input plane on which the conditional output p.m. $p_{Y|X}(y|x)$ is analytic (hence continuous) in the input variable.

The fundamental conditions on which our analysis relies may be summarized as follows:

- 1) The conditional output p.m. can be analytically extended to complex inputs, i.e., there exists an open set $\mathbb{A} \subseteq \mathbb{C}$ such that

$$x \mapsto p_{Y|X}(y|x)$$

is an analytic map over \mathbb{A} , while

$$(x, y) \mapsto p_{Y|X}(y|x)$$

is a continuous function over $\mathbb{A} \times \mathbb{R}$.

- 2) There exist two functions $q(y)$ and $Q(y)$, both nonnegative, and bounded above, and integrable, such that $\forall x \in \mathbb{S}$ we have

$$0 \leq q(y) \leq p_{Y|X}(y|x) \leq Q(y) \leq K < +\infty, \forall y \in \mathbb{R} \quad (8)$$

and the map

$$y \mapsto Q(y) \log q(y)$$

is integrable in y .

- 3) The two integrals

$$\begin{aligned} & \int_{\mathbb{R}} p_{Y|X}(y|w) \log p_{Y|X}(y|w) dy \\ & \int_{\mathbb{R}} p_{Y|X}(y|w) \log p_Y(y; p_X) dy \end{aligned}$$

are uniformly convergent (see [11] for definition) $\forall w \in \mathbb{D}$, for some \mathbb{D} such that $\mathbb{S} \subset \mathbb{D} \subseteq \mathbb{A}$.³

- 4) For each of the maximally extended connected regions forming \mathbb{S} (we call them S', S'', \dots), one of the following three conditions holds:
- a) there exist $x', x'', \dots \in S', S'', \dots$ (see Fig. 1) and corresponding c.d.f.'s F', F'', \dots with

$$\log p_{Y|X}(y|x') - \log q(y) < I(F'), \forall y \in \mathbb{R} \quad (9)$$

and analogously for the other regions, where $I(F')$ is the mutual information between the output and input variable when the input is distributed according to $F'(x)$.

- b) for all real input p.m.'s $p_X(x)$, there exist $x', x'', \dots \in S', S'', \dots$ (see Fig. 1) such that $p_{Y|X}(y|x')$ is the unique conditional output p.m. satisfying

$$\min_{x \in S'} D_{KL}(p_{Y|X}(y|x) || p_Y(y; p_X)) = D_{KL}(p_{Y|X}(y|x') || p_Y(y; p_X))$$

and analogously for the other regions, where D_{KL} denotes the Kullback-Leibler divergence.

³For the sake of clarity, here and elsewhere in the paper a generic input value is denoted by x or w whenever the input is considered strictly real or complex extended, respectively.

- c) for all real input p.m.'s $p_X(x)$, there exist pairs of distinct points $(x'_1, x'_2), (x''_1, x''_2), \dots \in S' \times S', S'' \times S'', \dots$ such that

$$D_{KL}(p_{Y|X}(y|x'_1)||p_Y(y;p_X)) \neq D_{KL}(p_{Y|X}(y|x'_2)||p_Y(y;p_X))$$

and analogously for the other regions.

Remark 1: The here stated conditions do not impose any peculiar kind of conditional output p.m., as it was the case in [1], [2], [7], nor any particular channel law, as it was done in [6]. We also underline that the input set compactness, deeply exploited in [7], is *not* a required condition here. Examples, considered in Section VIII, further show the presented theory to extend the previously known treatments.

We are now in a position to state the main result of this contribution.

Theorem 1: Every real scalar and peak power constrained input channel, whose conditional output p.m. fulfils the aforementioned conditions 1 to 4, has a finitely discrete capacity achieving input p.m.

The remainder of this paper is devoted to prove Theorem 1. The proof requires some intermediate steps: In particular, Section IV proves that the capacity achieving input p.m. exists and also states, as a corollary, Kuhn-Tucker's conditions on the marginal information density (defined in Section II) for an input p.m. to be optimal. Section V proves the analyticity of the marginal information density which is exploited in Section VI, alongside the corollary statement, to finally prove the finitely discrete nature of the capacity achieving input p.m. support. Besides, Section VII hints in the direction of proving uniqueness of the optimal input p.m.⁴

IV. EXISTENCE OF A CAPACITY ACHIEVING INPUT P.M.

Following the approach in [1], [2], in this section we demonstrate that an *optimal* input p.m. exists and that Kuhn-Tucker's conditions are necessary and sufficient for optimality. Some basic results in optimization theory are first reviewed [1], [2], [12].

A map $f : \Omega \mapsto \mathbb{R}$, where Ω is a convex space, is said to be weakly differentiable in Ω if, for $\theta \in [0, 1]$ and $x_0 \in \Omega$, the map $f'_{x_0} : \Omega \rightarrow \mathbb{R}$, defined as

$$f'_{x_0}(x) = \lim_{\theta \rightarrow 0} \frac{f[(1 - \theta)x_0 + \theta x] - f(x_0)}{\theta}$$

⁴Uniqueness was not proved in general neither in [6] nor in [7].

exists for all x and x_0 in Ω . Besides, f is said to be concave if, for all $\theta \in [0, 1]$ and for all x and x_0 in Ω ,

$$f[(1 - \theta)x_0 + \theta x] \geq (1 - \theta)f(x_0) + \theta f(x).$$

Theorem 2 (Optimization Theorem [12]): Let f be a continuous, weakly differentiable, and concave map from a compact, convex topological space Ω to \mathbb{R} , and define

$$C \triangleq \sup_{x \in \Omega} f(x).$$

Then:

- 1) $C = \max_{x \in \Omega} f(x) = f(x_0)$ for some $x_0 \in \Omega$;
- 2) $f(x_0) = C$ if and only if $f'_{x_0}(x) \leq 0 \ \forall x \in \Omega$.

Exploiting the above results from optimization theory, we have the following proposition.

Proposition 1: Let $I(F)$ be the mutual information functional between X and Y , as defined in (2). Then, under an input peak power constraint and conditions 1 and 2 of Sec. III, there exists an $F_0 \in \mathcal{F}$ (equivalently a $p_{X_0} \in \mathcal{P}$) such that

$$C = I(F_0) = \max_{F \in \mathcal{F}} I(F).$$

Moreover, a necessary and sufficient condition for the input c.d.f. F_0 to maximize $I(F)$, i.e., to achieve capacity, is

$$\int_{\mathbb{S}} i(x; F_0) dF(x) \leq I(F_0), \quad \forall F \in \mathcal{F}. \quad (10)$$

Proof: As from Theorem 2, it suffices to show that \mathcal{F} is convex and compact in some topology and that $I : \mathcal{F} \mapsto \mathbb{R}$ is continuous, concave and weakly differentiable. The necessary and sufficient condition (10) also follows from Theorem 2, as it will be shown.

a) Convexity and Compactness: The convexity of \mathcal{F} , i.e. the fact that

$$F_\theta(x) = (1 - \theta)F_1(x) + \theta F_2(x)$$

still belongs to \mathcal{F} for each F_1, F_2 in \mathcal{F} and for each $\theta \in [0, 1]$, is immediate. The compactness of \mathcal{F} in the Lèvy metric⁵ topology (as defined in [1]) follows from Helly's Weak Compactness Theorem (see Appendix D) and from the fact that convergence in the Lèvy metric is equivalent to complete convergence [13], which on a bounded interval is equivalent to weak convergence.

⁵The corresponding distance is here indicated with $d(\cdot, \cdot)$.

b) Continuity: The continuity of functional $I(F)$ descends from the Helly-Bray Theorem (see Appendix D), according to which $d(F_n, F) \xrightarrow{n} 0$ implies $I(F_n) \xrightarrow{n} I(F)$, provided the boundedness and continuity in x of $i(x; F)$. The latter two properties are demonstrated in Appendix A (continuity of $i(x; F)$ is a consequence also of analyticity discussed in Section V).

c) Concavity: For what concerns $I(F)$ being concave, we can note how

$$\begin{aligned} p_Y(y; F_\theta) &= p_Y(y; (1 - \theta)F_1 + \theta F_2) \\ &= \int_{\mathbb{S}} p_{Y|X}(y|x) [(1 - \theta)dF_1(x) + \theta dF_2(x)] = (1 - \theta)p_Y(y; F_1) + \theta p_Y(y; F_2) \end{aligned}$$

and

$$\begin{aligned} D((1 - \theta)F_1 + \theta F_2) &= - \int_{\mathbb{R}} \int_{\mathbb{S}} p(y|x) \log p(y|x) [(1 - \theta)dF_1(x) + \theta dF_2(x)] dy \\ &= (1 - \theta)D(F_1) + \theta D(F_2). \end{aligned} \tag{11}$$

Hence, we have that

$$I((1 - \theta)F_1 + \theta F_2) \geq (1 - \theta)I(F_1) + \theta I(F_2)$$

is equivalent, from (3) and (11), to

$$H((1 - \theta)F_1 + \theta F_2) \geq (1 - \theta)H(F_1) + \theta H(F_2). \tag{12}$$

Inequality (12) may be proved as follows:

$$\begin{aligned} H((1 - \theta)F_1 + \theta F_2) &= - \int_{\mathbb{R}} p_Y(y; (1 - \theta)F_1 + \theta F_2) \log p_Y(y; F_\theta) dy \\ &= - \int_{\mathbb{R}} [(1 - \theta)p_Y(y; F_1) + \theta p_Y(y; F_2)] \log p_Y(y; F_\theta) dy \\ &\stackrel{(a)}{\geq} -(1 - \theta) \int_{\mathbb{R}} p_Y(y; F_1) \log p_Y(y; F_1) dy - \theta \int_{\mathbb{R}} p_Y(y; F_2) \log p_Y(y; F_2) dy \\ &= (1 - \theta)H(F_1) + \theta H(F_2) \end{aligned}$$

where (a) exploits Gibbs' inequality [10], which states that for any two random variables, Z_1 and Z_2 , we have

$$- \int_{\mathbb{R}} p_{Z_1}(z) \log p_{Z_1}(z) dz \leq - \int_{\mathbb{R}} p_{Z_1}(z) \log p_{Z_2}(z) dz$$

with equality if and only if

$$p_{Z_1}(z) = p_{Z_2}(z).$$

Hence, concavity of $I(\cdot)$ is proven and equality holds if and only if $p_Y(y; F_1) = p_Y(y; F_2)$.

d) *Weak Differentiability*: As proven in Appendix B, for arbitrary F_1 and F_2 in \mathcal{F} we have

$$\lim_{\theta \rightarrow 0} \frac{I((1-\theta)F_1 + \theta F_2) - I(F_1)}{\theta} = \int_{\mathbb{S}} i(x; F_1) dF_2(x) - I(F_1). \quad (13)$$

The proof of weak differentiability is completed by observing that $i(x; F)$ is finitely bounded (Appendix A), which guarantees the existence of the integral in the right-hand side of (13).

Since all hypotheses of Theorem 2 are satisfied, the optimal input p.m. exists in \mathcal{P} . Furthermore, from (13), it is immediate to derive the necessary and sufficient condition (10). ■

The following corollary of Proposition 1 states the Kuhn-Tucker's conditions that will be used in Section VI to prove the final result.

Corollary 1 (Kuhn-Tucker's Conditions): Let p_{X_0} be an arbitrary p.m. in \mathcal{P} . Let S_0 denote the set of mass points of p_{X_0} on \mathbb{S} .⁶ Then p_{X_0} is optimal if and only if

$$\begin{cases} i(x; p_{X_0}) \leq I(p_{X_0}), & \forall x \in \mathbb{S} \\ i(x; p_{X_0}) = I(p_{X_0}), & \forall x \in S_0 \end{cases}$$

Proof: Even if Proposition 1 requires a different demonstration, the here stated corollary can be proved in the same way as done in [1], [2]. ■

V. ANALYTICITY OF $i(w; p_X)$

In this section we prove that $i(x; p_X)$ can be analytically extended to $i(w; p_X)$, $\forall w \in \mathbb{D}$. This step is necessary as a starting point for the capacity achieving input p.m. characterization in Section VI.

First, we extend $i(x; p_X)$ to the analyticity region \mathbb{A} of $x \mapsto p_{Y|X}(y|x)$ as

$$i(w; p_X) \triangleq \int_{\mathbb{R}} p_{Y|X}(y|w) \log \frac{p_{Y|X}(y|w)}{p_Y(y; p_X)} dy$$

$\forall w \in \mathbb{A}$ where convergence holds.⁷ We now apply the Differentiation Lemma (see Appendix D, with $I = \mathbb{R}$, $U = \mathbb{D}$), to the functions

$$f_1(w, y) = p_{Y|X}(y|w) \log p_{Y|X}(y|w),$$

$$f_2(w, y) = p_{Y|X}(y|w) \log p_Y(y; p_X).$$

⁶The set S_0 is defined independently of the discreteness or continuity of the input p.m.

⁷Convergence is guaranteed inside \mathbb{S} , as proven in Appendix A.

The two functions are continuous (see Section III) over $\mathbb{D} \times \mathbb{R}$.⁸ Moreover, from conditions in Section III, they are uniformly integrable over \mathbb{R} and, being compositions of analytic functions, they are analytic. The difference of the two analytic (from Differentiation Lemma) integral functions

$$\int_{\mathbb{R}} f_1(y, w) dy - \int_{\mathbb{R}} f_2(y, w) dy$$

is analytic on \mathbb{D} . This means that $i(w; p_X)$ is an analytic function over \mathbb{D} .

VI. CAPACITY ACHIEVING INPUT P.M. CHARACTERIZATION

In this section we finally prove the finite discreteness of the capacity achieving input p.m.. Define $v(w)$ as⁹

$$\begin{aligned} v(w) &\triangleq \int_{\mathbb{R}} p_{Y|X}(y|w) \left[-\log \left(\frac{p_Y(y; p_{X_0})}{p_{Y|X}(y|w)} \right) - I(p_{X_0}) \right] dy \\ &= i(w; p_{X_0}) - I(p_{X_0}) \end{aligned} \quad (14)$$

where $p_{X_0}(x)$ is a capacity achieving input p.m. Recall from Section III that S', S'', \dots are the maximally extended connected regions forming \mathbb{S} , while S'_0, S''_0, \dots is the corresponding decomposition for S_0 (the support of $p_{X_0}(x)$), i.e., S'_0 is the set of points of S_0 in S' , S''_0 is the set of points of S_0 in S'' , and so on. Note that, if each of the optimal input domain decomposition sets were not finitely discrete, then, for the Bolzano-Weierstrass Theorem, it would have an accumulation point in the corresponding connected subregion of \mathbb{S} and thus, by the identity principle of analytic functions and Corollary 1, $v(w) = 0$ in that entire subregion. From (14), $v(w) = 0$ means

$$-\int_{\mathbb{R}} p_{Y|X}(y|w) \log \left(\frac{p_Y(y; p_{X_0})}{p_{Y|X}(y|w)} \right) dy - I(p_{X_0}) = 0.$$

In the following, for notation convenience, suppose to consider the S' subregion of \mathbb{S} .

⁸ \mathbb{D} has to exclude the possibility for $p_{Y|X}(y|w)$ to be real negative valued, this to ensure continuity of the principal value complex logarithm.

⁹Recall that a generic input value is denoted by x and w when the input is considered strictly real or complex extended, respectively.

In case one of the first two options 4a, 4b presented in Section III is verified and since $v(w) = 0$ on the entire considered subregion, we must have:

$$v(x') = - \int_{\mathbb{R}} p_{Y|X}(y|x') \log \left(\frac{p_Y(y; p_{X_0})}{p_{Y|X}(y|x')} \right) dy - I(p_{X_0}) = 0$$

also for the corresponding particular value x' , whose existence was supposed in Section III. However this is in clear contradiction with either

$$\begin{aligned} v(x') &= \int_{\mathbb{R}} p_{Y|X}(y|x') \log \left(\frac{p_{Y|X}(y|x')}{p_Y(y; p_{X_0})} \right) dy - I(p_{X_0}) \\ &\leq \int_{\mathbb{R}} p_{Y|X}(y|x') \underbrace{\left(\log p_{Y|X}(y|x') - \log q(y) \right)}_{< I(F') \text{ see eq.(9)} } dy - I(p_{X_0}) < I(F') - I(F_0) \leq 0. \end{aligned}$$

or

$$\begin{aligned} v(x') &= \int_{\mathbb{R}} p_{Y|X}(y|x') \log \left(\frac{p_{Y|X}(y|x')}{p_Y(y; p_{X_0})} \right) dy - I(p_{X_0}) \\ &= D_{KL}(p_{Y|X}(y|x') || p_Y(y; p_{X_0})) - I(p_{X_0}) < D_{KL}(p_{Y|X}(y|x) || p_Y(y; p_{X_0})) - I(p_{X_0}) = 0. \end{aligned}$$

If vice versa the third option 4c holds, it follows

$$v(x'_1) = D_{KL}(p_{Y|X}(y|x'_1) || p_Y(y; p_{X_0})) - I(p_{X_0}) \neq D_{KL}(p_{Y|X}(y|x'_2) || p_Y(y; p_{X_0})) - I(p_{X_0}) = 0$$

and again a contradiction occurs.

This finally proves that the hypothesis to have an infinite set of mass points S_0 was wrong, hence the input RV X can take only on a finitely discrete set of values.

VII. ABOUT UNIQUENESS

The so far developed conditions on the capacity achieving input p.m. do not guarantee also its uniqueness. In this direction, a further property that all eventual optimal input p.m.'s must satisfy with respect to any other capacity achieving p.m. can be outlined.

Consider all the optimal input p.m.'s¹⁰ and denote the i -th of them by $p_{X_i}(x)$. Then, the following proposition holds.

¹⁰In the previous sections, we proved that they belong to \mathcal{P}' , the restriction of \mathcal{P} to the class of finitely discrete generalized functions defined on a finite number of probability mass points in the input support \mathbb{S} .

Proposition 2: All the optimal input p.m.'s of a channel model satisfying conditions 1-4 in Section III, must fulfil the condition

$$i(x; p_{X_0}) = I(p_{X_0}), \quad \forall x \in S_i$$

$S_i \subset \mathbb{S}$ being the support of $p_{X_i}(x)$.

Proof: Let $p_{X_0}(x)$ and $p_{X_1}(x)$ be two optimal input p.m.'s (whose existence is guaranteed by Proposition 1), both with a finitely discrete support. Then also $(1-\theta)p_{X_1}(x) + \theta p_{X_0}(x)$ is capacity achieving, since the mutual information functional is concave (see Theorem 6 in Appendix D). This fact yields the weak derivative $I'_{p_{X_0}}(p_{X_1})$ to be null. Recall the probability mass points in S_0 and S_1 x_m and x_n , and the correspondent probability b_m and a_n , respectively. In addition suppose that the condition enunciated in Proposition 2 is not verified, i.e., $i(x; p_{X_0}) < I(p_{X_0})$ for at least one of the $x_n \in S_1$, where the order relation is imposed by Corollary 1. The cited weak derivative expression becomes

$$\begin{aligned} \int_{\mathbb{S}} i(x; p_{X_0}) [p_{X_1}(x) - p_{X_0}(x)] dx &= \sum_n a_n i(x_n; p_{X_0}) - \sum_m b_m i(x_m; p_{X_0}) \\ &< I(p_{X_0}) \sum_n a_n - I(p_{X_0}) \sum_m b_m = 0. \end{aligned}$$

A contradiction has arisen since $I'_{p_{X_0}}(p_{X_1}) = 0$ and $I'_{p_{X_0}}(p_{X_1}) < 0$, which completes the proof. ■

This Proposition 2 does not provide uniqueness of the capacity achieving input p.m., nevertheless it tightens the conditions for an input p.m. to be optimal. Future attempts will be made aiming to prove uniqueness.

VIII. EXAMPLES

This section is divided in two subsections. The first one proves that any peak power constrained channel with additive noise satisfies condition 4, stated in Section III and, therefore, it belongs to the general class of channels treated in this paper upon fulfilling also conditions 1, 2, and 3.¹¹ The second one proves that the Rayleigh fading channel undergoes all the conditions in Section III. With respect to the theory proposed in [6], we underline that the conditions in Section III are less stringent, so a wider set of additive channels is characterized.

¹¹The fulfilment of conditions 1-3 must be checked case by case, but it is expected to be a simple verification.

A. Additive Channels

Consider an additive channel model $Y = X + N$, where N is the noise RV. The marginal information density can be rewritten as

$$\begin{aligned} i(x; p_X) &= \int_{\mathbb{R}} p_N(y - x) \log p_N(y - x) dy - \int_{\mathbb{R}} p_N(y - x) \log p_Y(y; p_X) dy \\ &= k - \int_{\mathbb{R}} p_N(y - x) \log p_Y(y; p_X) dy \end{aligned}$$

where k is constant as it can be easily shown with an ordinary variable substitution. The second term is in the form of convolution and admits Fourier Transform (FT) since $p_N(\cdot)$ is integrable on \mathbb{R} and $\log p_Y(y; p_X) = u(y)$ is locally integrable hence transformable at least in the sense of distributions. Now assume the marginal information density is equal to a constant c_1 : Its FT would then be

$$\Psi_N(2\pi f)U(f) = c_1\delta(f)$$

where $\Psi_N(\cdot)$ denotes the characteristic function of the RV N , defined as

$$\Psi_N(f) = \mathbb{E}[\exp\{jxf\}] = \int_{\mathbb{R}} p_N(x) \exp\{jxf\} dx.$$

The only case for this to hold is $u(y)$ being a constant itself: This is however contradictory since $u(y) = c_2$ implies $p_Y(y; p_X) = 2^{c_2}$, which is clearly an absurd, and hence condition 4c stands.

B. Rayleigh Fading Channel

Consider the Rayleigh fading channel conditional output p.m., as defined in [5],

$$\begin{aligned} p_{Y|X}(y|x) &= \frac{1}{1+x^2} \exp\left\{-\frac{y}{1+x^2}\right\} \\ &= s \exp\{-ys\} \end{aligned}$$

and assume the channel input X is subject to a peak power constraint A as defined in Section III. Since this conditional p.m. derives from normalizations of the original input and output modules, U and V in [5], this is a real scalar memoryless channel whose output takes values in $[0, +\infty)$. We now assess that the four conditions stated in Section III are fulfilled.

- (a) It is immediate to verify that condition 1 holds over the set $\mathbb{A} = \mathbb{C} \setminus \{-j, j\}$.

(b) Concerning condition 2, let us define

$$Q(y) = \begin{cases} 1, & 0 \leq y \leq c(A^2 + 1) \\ \frac{1}{y^{1+\gamma}}, & y > c(A^2 + 1) \end{cases}$$

where parameter γ fulfils $\gamma < 1$ and c is a constant such that $c > 2$ (the details are provided in Appendix C). Moreover, let us define

$$q(y) = \begin{cases} \frac{1}{1+A^2} \exp\left\{-\frac{y}{1+A^2}\right\}, & 0 \leq y \leq \frac{(1+A^2)\ln(1+A^2)}{A^2} \\ \exp\{-y\}, & y > \frac{(1+A^2)\ln(1+A^2)}{A^2} \end{cases}$$

where $y_2 = \frac{(1+A^2)\ln(1+A^2)}{A^2}$ is the solution of $\frac{1}{1+A^2} \exp\left\{-\frac{y}{1+A^2}\right\} = \exp\{-y\}$. The two functions $q(\cdot)$ and $Q(\cdot)$ satisfy inequality (8), as rigorously proven in Appendix C. Furthermore, both of them are nonnegative, superiorly bounded, and integrable over the output domain $[0, +\infty)$. Besides $Q(y) \log q(y)$ is integrable over $[0, +\infty)$, which may be shown by analysing integrability over the tail.¹² We have

$$\begin{aligned} \int_{y_3}^{+\infty} Q(y) \log q(y) dy &= \int_{y_3}^{+\infty} \frac{1}{y^{1+\gamma}} \exp\{-y\} dy \\ &= \left[-\frac{y^{-\gamma}}{\gamma} \exp\{-y\} \right]_{y_3}^{+\infty} - \int_{y_3}^{+\infty} \frac{y^{-\gamma}}{\gamma} \exp\{-y\} dy \end{aligned}$$

which is finite. The considered y_3 is sufficiently large to guarantee that the expressions employed for $Q(y)$ and $q(y)$ are the proper ones.

(c) We now consider condition 3. The integral

$$\int_0^{+\infty} p_{Y|X}(y|w) \log p_Y(y; p_X) dy$$

is uniformly convergent on $\mathbb{D} = \{w : \Re\{\frac{1}{1+w^2}\} \geq a_1, |\frac{1}{1+w^2}| \leq a_2\}$, with strictly positive a_1 and a_2 , and with a_1 ensuring that $\mathbb{S} \subset \mathbb{D}$. Uniform convergence holds since, for each

¹² $Q(y) \log q(y)$ is locally integrable since it is continuous.

$w \in \mathbb{D}$, given ϵ , there exist $B_0 < B_1 < B_2$ such that

$$\begin{aligned}
\left| \int_{B_1}^{B_2} p_{Y|X}(y|w) \log p_Y(y; p_X) dy \right| &\leq \int_{B_1}^{B_2} |p_{Y|X}(y|w) \log p_Y(y; p_X)| dy \\
&= \int_{B_1}^{B_2} \left| \frac{1}{1+w^2} \right| \left| \exp \left\{ -\frac{y}{1+w^2} \right\} \log p_Y(y; p_X) \right| dy \\
&= \int_{B_1}^{B_2} \left| \frac{1}{1+w^2} \right| \left| \exp \left\{ -y \Re \left\{ \frac{1}{1+w^2} \right\} \right\} \log p_Y(y; p_X) \right| dy \\
&\leq - \int_{B_1}^{B_2} \frac{1}{y^3} \log q(y) dy < \epsilon
\end{aligned}$$

as $\left| \frac{1}{1+w^2} \exp \left\{ -y \Re \left\{ \frac{1}{1+w^2} \right\} \right\} \right|$ is minor in a definitive manner in y than $1/y^3$ regardless of $w \in \mathbb{D}$.¹³ To prove the result it is also necessary to employ (18) in Appendix A and to choose B_0 in such a way that $B_0 > y_2$ and $\frac{1}{B_0} < \epsilon$. The choice for \mathbb{D} is dictated by the necessity to guarantee the existence of a uniform upper bound for $|p_{Y|X}(y|w) \log p_Y(y; p_X)|$. Analogously, also

$$\int_0^{+\infty} p_{Y|X}(y|w) \log p_{Y|X}(y|w) dy$$

is uniformly convergent on \mathbb{D} . In fact, for each $w \in \mathbb{D}$, given ϵ , there exist $B_0 < B_1 < B_2$ such that

$$\begin{aligned}
\left| \int_{B_1}^{B_2} p_{Y|X}(y|w) \log p_{Y|X}(y|w) dy \right| &\leq \int_{B_1}^{B_2} |p_{Y|X}(y|w) \log p_{Y|X}(y|w)| dy \\
&\leq \int_{B_1}^{B_2} \left| \frac{1}{1+w^2} \right| \left| \exp \left\{ -\frac{y}{1+w^2} \right\} \log \frac{1}{1+w^2} \right| dy \\
&\quad + \int_{B_1}^{B_2} \left| \frac{1}{1+w^2} \right| \left| \exp \left\{ -\frac{y}{1+w^2} \right\} \log \left(\exp \left\{ -\frac{y}{1+w^2} \right\} \right) \right| dy \\
&\leq \int_{B_1}^{B_2} \frac{1}{y^2} dy < \epsilon
\end{aligned}$$

where again B_0 is chosen to ensure $1/B_0 < \epsilon$.

(d) We finally have to address condition 4. Consider

$$\begin{aligned}
&\int_0^{+\infty} p_{Y|X}(y|x) \log p_{Y|X}(y|x) dy - \int_0^{+\infty} p_{Y|X}(y|x) \log p_Y(y; p_X) dy \\
&= \int_0^{+\infty} s \exp\{-ys\} \log(s \exp\{-ys\}) dy - \int_0^{+\infty} s \exp\{-ys\} \log p_Y(y; p_X) dy \\
&= \log s - \frac{1}{\ln 2} - \int_0^{+\infty} s \exp\{-ys\} \log p_Y(y; p_X) dy.
\end{aligned} \tag{15}$$

¹³This is guaranteed by the existence of a maximum for $|\frac{1}{1+w^2}|$ and a non zero minimum for $\Re \left\{ \frac{1}{1+w^2} \right\}$ on \mathbb{D} .

The third term dependence¹⁴ on s cannot be logarithmic since

$$\lim_{s \rightarrow +\infty} - \int_0^{+\infty} s \exp\{-ys\} \log p_Y(y; p_X) dy = 0$$

where exchange between integral and limit is licit since when $s \rightarrow +\infty$ it can be supposed greater than 1, this ensuring the existence of an integrable upper bound of $|s \exp\{-ys\} \log p_Y(y; p_X)|$, much as previously done for integrability of $Q(y) \log q(y)$. Hence the difference between the first and third term of (15) cannot be constant on \mathbb{S} , this proving condition 4c to hold.

IX. CONCLUSION

This paper has proposed general conditions on the conditional output p.m. under which real scalar channel models, with input peak power constraints, show to have capacity achieving p.m.'s which are finitely discrete. These conditions represent a step towards a full understanding of the basic channel characteristics that determine the capacity achieving input p.m. to be finitely discrete under peak power constraints. The here presented theory of peak power limited channels unifies under a same framework several channel models that were previously investigated using separated approaches, as shown by the provided examples.

Particular attention will be paid in the future to whether all of the supposed conditions are strictly necessary. Our feeling is that some of those conditions are not negotiable while other ones may not be as fundamental as they can appear to be.

As last but not least consideration, we have matured the deep belief that only *real* scalar peak power limited channels can have a finitely discrete capacity achieving input probability measure.

APPENDIX A

BOUNDEDNESS AND CONTINUITY OF THE MARGINAL INFORMATION DENSITY

The existence and boundedness of the upper and lower bounds on $p_{Y|X}(y|x)$, postulated in Section III is sufficient to prove the existence and boundedness of $p_Y(y; p_X)$. In fact, we can write

$$q(y) = \int_{\mathbb{S}} q(y) p_X(x) dx \leq \int_{\mathbb{S}} p_{Y|X}(y|x) p_X(x) dx \leq \int_{\mathbb{S}} Q(y) p_X(x) dx = Q(y)$$

¹⁴Dependence on variable s is the same independently of the s considered: It is thus possible to consider values for s even outside the region dictated by the particular channel capacity problem we are considering.

that is

$$0 \leq q(y) \leq p_Y(y; p_X) \leq Q(y) \leq K, \quad \forall y \in \mathbb{R} \text{ and } \forall p_X(x) \in \mathcal{P}. \quad (16)$$

An equally useful inequality, immediately descending from the previous one, is the following:

$$-\log Q(y) \leq -\log p_Y(y; p_X) \leq -\log q(y), \quad \forall y \in \mathbb{R} \text{ and } \forall p_X(x) \in \mathcal{P}. \quad (17)$$

Moreover, consider the pair of functions $f(y)$ and $g(y)$, respectively nonnegative and positive, such that $g(y) \leq K < +\infty$. The next inequality holds:

$$\begin{aligned} |f(y) \log g(y)| &\leq -f(y) \log \frac{g(y)}{K} + f(y) |\log K| \\ &\leq -f(y) \log g(y) + 2f(y) |\log K|. \end{aligned} \quad (18)$$

Besides

$$G(y) = -Q(y) \log q(y) + 2Q(y) |\log K|$$

is integrable on \mathbb{R} . Proof for this is an immediate consequence of the conditions in Sec. III.

We now show that $h(x; p_X)$ and $i(x; p_X)$ are bounded $\forall x \in \mathbb{S}$ and $\forall p_X(x) \in \mathcal{P}$. In fact we have

$$\begin{aligned} |h(x; p_X)| &= \left| \int_{\mathbb{R}} p_{Y|X}(y|x) \log p_Y(y; p_X) dy \right| \leq \int_{\mathbb{R}} |p_{Y|X}(y|x) \log p_Y(y; p_X)| dy \\ &\leq \int_{\mathbb{R}} p_{Y|X}(y|x) \left[-\log p_Y(y; p_X) + 2|\log K| \right] dy \leq \int_{\mathbb{R}} Q(y) [-\log q(y) + 2|\log K|] dy \\ &= \int_{\mathbb{R}} G(y) dy < +\infty \end{aligned}$$

having used (8), (16), (17) and (18). Moreover, we have

$$\begin{aligned} |d(x)| &= \left| - \int_{\mathbb{R}} p_{Y|X}(y|x) \log p_{Y|X}(y|x) dy \right| \leq \int_{\mathbb{R}} |p_{Y|X}(y|x) \log p_{Y|X}(y|x)| dy \\ &\leq \int_{\mathbb{R}} p_{Y|X}(y|x) \left[-\log p_{Y|X}(y|x) + 2|\log K| \right] dy \leq \int_{\mathbb{R}} Q(y) [-\log q(y) + 2|\log K|] dy \\ &= \int_{\mathbb{R}} G(y) dy < +\infty \end{aligned}$$

where we again exploited (8), (16), (17) and (18). We may then conclude that $i(x; p_X) = h(x; p_X) - d(x)$ is bounded, as it is the difference between two quantities fulfilling the same finite boundedness property.

Continuity of $i(x; p_X)$ can be demonstrated in an almost identical way since, $\forall x \in \mathbb{S}$, it is possible to exchange the continuity limit with the integral in the definition of $i(\cdot)$, this being guaranteed by integrability of $G(y)$, and continuity of the integrand functions being an immediate evidence.

APPENDIX B

PROOF OF EQUATION (13)

The weak derivative can be developed as

$$\begin{aligned}
& \lim_{\theta \rightarrow 0} \frac{I((1-\theta)F_1 + \theta F_2) - I(F_1)}{\theta} \\
&= \lim_{\theta \rightarrow 0} \left\{ \frac{1}{\theta} \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{p(y; (1-\theta)F_1 + \theta F_2)} [(1-\theta)dF_1(x) + \theta dF_2(x)] dy \right. \\
&\quad \left. - \frac{1}{\theta} \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{p_Y(y; F_1)} dF_1(x) dy \right\} \\
&= \lim_{\theta \rightarrow 0} \left\{ \frac{1}{\theta} \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) [-\log p_Y(y; F_\theta) + \log p_Y(y; F_1)] dF_1(x) dy \right. \\
&\quad \left. + \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{(1-\theta)p_Y(y; F_1) + \theta p_Y(y; F_2)} [dF_2(x) - dF_1(x)] dy \right\} \\
&\stackrel{(a)}{=} \lim_{\theta \rightarrow 0} \left\{ \frac{1}{\theta} \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) [\log p_Y(y; F_1) - \log ((1-\theta)p_Y(y; F_1) + \theta p_Y(y; F_2))] dF_1(x) dy \right\} \\
&\quad + \int_{\mathbb{S}} i(x; F_1) dF_2(x) - I(F_1) \\
&\stackrel{(b)}{=} \lim_{\theta \rightarrow 0} \left\{ \frac{1}{\theta} \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \left[\log p_Y(y; F_1) - \log p_Y(y; F_1) \right. \right. \\
&\quad \left. \left. - \frac{1}{\ln 2} \theta \frac{-p_Y(y; F_1) + p_Y(y; F_2)}{p_Y(y; F_1)} \right] dF_1(x) dy \right\} + \int_{\mathbb{S}} i(x; F_1) dF_2(x) - I(F_1) \\
&= \frac{1}{\ln 2} \int_{\mathbb{R}} \int_{\mathbb{S}} p_{Y|X}(y|x) \left(1 - \frac{p_Y(y; F_2)}{p_Y(y; F_1)} \right) dF_1(x) dy + \int_{\mathbb{S}} i(x; F_1) dF_2(x) - I(F_1) \\
&= \int_{\mathbb{S}} i(x; F_1) dF_2(x) - I(F_1) + \frac{1}{\ln 2} \int_{\mathbb{R}} p_Y(y; F_1) \left(1 - \frac{p_Y(y; F_2)}{p_Y(y; F_1)} \right) dy \\
&= \int_{\mathbb{S}} i(x; F_1) dF_2(x) - I(F_1)
\end{aligned}$$

where the exchange between limit and integral in (a) follows from the Lebesgue Dominated Convergence Theorem. In fact, $\forall \theta \in [0, 1]$

$$\begin{aligned} f_\theta(y, x)p_{X_2}(x) &= p_{Y|X}(y|x)p_{X_2}(x) \log \frac{p_{Y|X}(y|x)}{p_Y(y; F_\theta)} \\ &\leq p_{X_2}(x) \left(|p_{Y|X}(y|x) \log p_{Y|X}(y|x)| + |p_{Y|X}(y|x) \log p_Y(y; F_\theta)| \right) \end{aligned}$$

which is integrable on $\mathbb{R} \times \mathbb{S}$,¹⁵ by integrability of $G(y)$, and then also on $\mathbb{S} \times \mathbb{R}$ via Tonelli and Fubini Theorems and due to the fact that $f_\theta(y, x)$ converges, for $\theta \rightarrow 0$, to $f(y, x) = p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{p_Y(y; F_1)}$. Moreover, (b) follows from the first order McLaurin Series

$$\log(a(1-x) + bx) = \log(a) + \frac{x(-a+b)}{a} \frac{1}{\ln 2} + o(x).$$

APPENDIX C

AN UPPER AND LOWER BOUND FOR THE RAYLEIGH FADING CONDITIONAL OUTPUT P.M.

In this appendix we rigorously prove inequality (8) to be satisfied in case the considered conditional output p.m. and correspondent $Q(y)$ and $q(y)$ are the ones introduced in Section VIII-B. Concerning the upper bound, we have to show that there exist a parameter γ such that

$$\frac{1}{y^{1+\gamma}} > \frac{1}{1+x^2} \exp \left\{ -\frac{y}{1+x^2} \right\}, \quad \forall x : 0 \leq |x| \leq A \quad (19)$$

is valid for $y > c(A^2 + 1)$, where $c > 2$. The considered inequality can be reformulated as follows

$$\frac{y}{(1+x^2)(1+\gamma)} + \frac{\ln(1+x^2)}{1+\gamma} > \ln y.$$

To guarantee the inequality to be fulfilled even in the worst case, the left hand side (x is confined in it) can be studied, for each fixed y , to find out that $\sqrt{y-1}$ is its minimum in x , provided $y \geq 1$. Moreover, if $y \geq A^2 + 1$ the minimum becomes $x = A$, since x is bounded and $\sqrt{y-1}$ is unreachable in this case. The minimum expression for $y \geq A^2 + 1$ is

$$\frac{y}{(1+A^2)(1+\gamma)} + \frac{\ln(1+A^2)}{1+\gamma}$$

¹⁵Integration on \mathbb{R} produces $p_{X_2}(x)i(x; F_\theta)$ that is integrable on \mathbb{S} due to the boundedness of $i(\cdot)$.

which has constant derivative in y equalling the derivative of $\ln y$ at $y = (1 + A^2)(1 + \gamma)$. Now, consider $y_1 = c(1 + A^2)$: If constants c and γ are chosen such that $c > 2$ and $\gamma < 1$, then $y_1 > (1 + A^2)(1 + \gamma)$ and

$$\left. \frac{d}{dy} \left[\frac{y}{(1 + A^2)(1 + \gamma)} + \frac{\ln(1 + A^2)}{1 + \gamma} \right] \right|_{y_1} > \left. \frac{d}{dy} \ln y \right|_{y_1}.$$

This ensures the derivative of $\frac{y}{(1 + A^2)(1 + \gamma)} + \frac{\ln(1 + A^2)}{1 + \gamma}$ to be greater than the one of $\ln y$, which is decreasing, for $y > y_1$. If, finally, it is possible to derive a condition on γ to provide that

$$\left(\frac{y}{(1 + A^2)(1 + \gamma)} + \frac{\ln(1 + A^2)}{1 + \gamma} \right) \Big|_{y_1} > \ln y|_{y_1} \quad (20)$$

the original assertion (19) would be satisfied. This is indeed possible since (20) becomes

$$\frac{c}{1 + \gamma} + \frac{\ln(1 + A^2)}{1 + \gamma} > \ln(1 + A^2) + \ln c$$

which is satisfied for $\gamma < \frac{c - \ln c}{\ln[c(1 + A^2)]}$. Any choice of γ such that

$$\gamma < \min \left\{ 1, \frac{c - \ln c}{\ln[c(1 + A^2)]} \right\}$$

would fulfil the scope. Consequently the definition

$$Q(y) = \begin{cases} 1, & 0 \leq y \leq c(A^2 + 1) \\ \frac{1}{y^{1+\gamma}}, & y > c(A^2 + 1) \end{cases}$$

is well posed since it guarantees the right hand side of inequality (8) to be respected.

Concerning the lower bound $q(y)$, we have to prove that it coincides with the output p.m. conditioned by the maximum input up to $y_2 = \frac{(1 + A^2)\ln(1 + A^2)}{A^2}$ and that it coincides with the output p.m. conditioned by the minimum input after that same y_2 . To do that, consider the intersection between $\frac{1}{1+x^2} \exp\left\{-\frac{y}{1+x^2}\right\}$ and $\exp\{-y\}$ which is given by

$$y(x) = \frac{(1 + x^2) \ln(1 + x^2)}{x^2}.$$

This intersection is non decreasing in x for $0 \leq x \leq A$ (only positive values are admissible for x , deriving from normalization in [5]) since

$$\frac{dy(x)}{dx} = \frac{2}{x^3} [x^2 - \ln(1 + x^2)] \geq 0$$

this meaning that it is maximum for $x = A$. This duly proves that the output p.m. conditioned by the maximum input lies under all the other conditional output p.m. up to its intersection with

$\exp\{-y\}$ in $y_2 = \frac{(1+A^2)\ln(1+A^2)}{A^2}$, while afterwards the same role is taken by $\exp\{-y\}$. This finally proves that

$$q(y) = \begin{cases} \frac{1}{1+A^2} \exp\left\{-\frac{y}{1+A^2}\right\}, & 0 \leq y \leq \frac{(1+A^2)\ln(1+A^2)}{A^2} \\ \exp\{-y\}, & y > \frac{(1+A^2)\ln(1+A^2)}{A^2} \end{cases}$$

is also well posed, fulfilling the left hand side of (8).

APPENDIX D USEFUL THEOREMS

This appendix provides a collection of theorem statements (along with the appropriate references) that are used throughout this paper.

Theorem 3 (Helly's Weak Compactness Theorem [14]): Every sequence of c.d.f.'s is weakly compact.¹⁶

Theorem 4 (Helly-Bray Theorem [14]): If g is continuous and bounded on \mathbb{R}^n , then $F_n \xrightarrow[n]{c} F$ up to additive constants implies $\int g dF_n \rightarrow \int g dF$.

This theorem is formulated in terms of complete convergence, but complete convergence is equivalent to Lèvy convergence in \mathcal{F} .

Theorem 5 (Differentiation Lemma [11]): Let I be an interval of real numbers, eventually infinite, and U be an open set of complex numbers. Let $f = f(t, z)$ be a continuous function on $I \times U$. Assume 1) for each compact subset \mathbb{K} of U the integral $\int_I f(t, z) dt$ is uniformly convergent for $z \in \mathbb{K}$, 2) for each t , the function $z \mapsto f(t, z)$ is analytic, then the integral function $F(z) = \int_I f(t, z) dt$ is analytic on U .

Theorem 6 ([12], Proposition 1, Chapter 7.8): Let f be a concave functional defined on a convex subset C of a normed space. Let $\mu = \sup_{x \in C} f(x)$. Then

- 1) The subset Ω of C where $f(x) = \mu$ is convex.
- 2) If x_0 is a local maximum of $f(\cdot)$, then $f(x_0) = \mu$ and, hence x_0 is a global maximum.

¹⁶Recall that a set is said to be *compact*, in the sense of a type of convergence, if every infinite sequence in the set contains a subsequence which is convergent in that same sense [14].

ACKNOWLEDGMENT

The authors would like to thank Prof. Marco Chiani, Prof. Massimo Cicognani, Andrea Mariani, Simone Moretti and Matteo Mazzotti for useful comments and discussions.

REFERENCES

- [1] J. G. Smith, “On the information capacity of peak and average power constrained Gaussian channels,” *Ph.D. dissertation, University of California, Berkeley*, 1969.
- [2] —, “The information capacity of amplitude- and variance-constrained scalar Gaussian channels,” *Information and Control*, vol. 18, 1971.
- [3] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell System Technical Journal*, Jul./Oct. 1948.
- [4] S. Shamai and I. Bar-David, “The Capacity of Average and Peak-Power-Limited Quadrature Gaussian Channels,” *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060–1071, Jul. 1995.
- [5] I. Abou-Faycal, M. Trott, and S. Shamai, “The Capacity of Discrete-Time Memoryless Rayleigh-Fading Channels,” *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.
- [6] A. Tchamkerten, “On the Discreteness of Capacity-Achieving Distributions,” *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2773–2778, Nov. 2004.
- [7] T. H. Chan, S. Hranilovic, and F. R. Kschischang, “Capacity-Achieving Probability Measure for Conditionally Gaussian Channels With Bounded Inputs,” *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, Jun. 2005.
- [8] A. Feiten and R. Mathar, “Capacity-Achieving Discrete Signaling over Additive Noise Channels,” in *Proc. IEEE Int. Conf. on Commun.*, Jun. 2007, pp. 5401–5405.
- [9] E. Leitinger, B. C. Geiger, and K. Witrisal, “Capacity and Capacity-Achieving Input Distribution of the Energy Detector,” in *IEEE Int. Conf. on Utra-Wideband*, Sep. 2012, pp. 57–61.
- [10] T. A. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York, NY, 10158: John Wiley & Sons, Inc., 1991.
- [11] S. Lang, *Complex Analysis*. New York: Springer-Verlag, 1999.
- [12] D. Luenberger, *Optimization by Vector Space Methods*. New York: John Wiley & Sons, 1969.
- [13] P. Moran, *An Introduction to Probability Theory*. Oxford: Clarendon Press, 1968.
- [14] M. Loève, *Probability Theory I*, 4th ed. New York, NY: Springer-Verlag, 1977.